ARI Research Note 88-45

DTIC FILE COPY

(1)

AD-A198 456

Base Rate Effects on the Interpretation of Probability and Frequency Expressions

Thomas S. Wallsten, Samuel Fillenbaum,
and James A. Cox
University of North Carolina

for

Contracting Officer's Representative
Michael Drillings

ARI Scientific Coordination Office, London
Milton S. Katz, Chief

Basic Research Laboratory
Michael Kaplan, Director

U. S. Army
Research Institute for the Behavioral and Social Sciences

June 1988

88 8

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

L. NEALE COSBY
Colonel, IN
Commander

EDGAR M. JOHNSON
Technical Director

Research accomplished under contract
for the Department of the Army

University of North Carolina

Technical review by

Dan Ragland

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Note 88-45	2. GOVT ACCESSION NO. - -	3. RECIPIENT'S CATALOG NUMBER - -
4. TITLE *(and Subtitle)* Base Rate Effects on the Interpretations of Probability and Frequency Expressions		5. TYPE OF REPORT & PERIOD COVERED Interim Report November 84 - November 85
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas S. Wallsten, Samuel Fillenbaum and James A. Cox		8. CONTRACT OR GRANT NUMBER(s) MDA903-83-K-0347
9. PERFORMING ORGANIZATION NAME AND ADDRESS L.L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill, NC 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS U.S, Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE June 1988
		13. NUMBER OF PAGES 43
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* - -		15. SECURITY CLASS. *(of this report)* Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE n/a

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

- -

18. SUPPLEMENTARY NOTES

Michael Drillings, contracting officer's representative

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Cognitive Science
Decision Making
Psychometrics
Judgement

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This research note reports on two studies run to determine whether the interpretations of statements, or forecasts, using vague probability and frequency expressions such as "likely", "improbable","frequently", "rarely" were sensitive to the base rates of the events involved. In the first experiment, professional weather forecasters judged situations drawn from a medical context. In the second, students judged matched forecast scenarios of common semantic content that differed only in prior probability (as determined by an independent group

(OVER)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

ARI RESEARCH NOTE 88-44

20. Abstract (continued)

of subjects). Results were as follows: the interpretations of forecasts using neutral terms (e.g. possible) and terms above neutral (e.g. usually) were strong, positive functions of base rate, while the interpretations of forecasts using terms below neutral (e.g. rarely) were much less affected by base rates; in the second experiment, interpretations of forecasts appeared to represent some kind of average of the meaning of the expression and the base rate.

Abstract

Two studies were run to determine whether the interpretations of statements or forecasts using vague probability and frequency expressions such as *likely*, *improbable*, *frequently*, *rarely*, were sensitive to the base rates of the events involved. In the first experiment, professional weather forecasters judged situations drawn from a medical context. In the second, students judged matched forecast scenarios of common semantic content that differed only in prior probability (as determined by an independent group of subjects). Results were: (a) The interpretations of forecasts using neutral terms (e.g., *possible*) and terms above neutral (e.g., *usually*) were strong, positive functions of base rate, while the interpretations of forecasts using terms below neutral (e.g., *rarely*) were much less affected by base rates; (b) In the second experiment interpretations of forecasts appeared to represent some kind of average of the meaning of the expression and the base rate.

The question of whether the meanings of nonnumerical expressions of uncertainty depend on context, and if so, how, is important for related practical and theoretical reasons. The practical issues arise from the fact that most people, including many expert forecasters, generally prefer communicating their uncertain opinions with vague expressions such as doubtful, probable, or unusual, rather than numerically. The theoretical issues arise, of course, from an attempt to understand how judgment is formed, modified, and communicated on the basis of such expressions.

On anecdotal grounds, people prefer the imprecision of nonnumerical phrases to the precision of numbers for at least two reasons. First, their opinions or judgments are generally not precise, and therefore it would be misleading to represent them precisely. Second, people feel that they better understand the meanings of words than of numbers, and therefore that their opinions are better conveyed verbally than numerically. This point has been made from a historical perspective by Zimmer (1984), who noted that verbal expressions of uncertainty were available long before the development of mathematical probability concepts in the 17th century. Zimmer further suggested that people process uncertainty in a verbal rather than a numerical manner and that judgments are revised in light of new information according to linguistic rather than numerical principles.

An important requirement for the effective use of vague expressions in communication is that their meanings be relatively constant over contexts. However, if Zimmer is correct that verbally stated uncertainties are processed linguistically, then it is doubtful that this requirement is met, because the meanings of words are frequently and systematically influenced by the contexts in which they are embedded (e.g., Kess & Hoppe, 1981, and the bibliography in Fries, 1980).

In many conversational situations, meaning is sensitive to context, but communication does not suffer, because speaker and listener share common assumptions and knowledge so that context effects are identical for both of them (e.g., Searle, 1975, and other essays in Cole & Morgan, 1975). However, it is particularly in situations of uncertainty that communicating parties are most likely to have different assumptions and knowledge, and therefore for context to differentially affect their understanding of words and expressions.

It is worth mentioning at this point that there have been recent suggestions within the context of fuzzy set theory that the vague meanings of probability or frequency expressions (or of linguistic variables more generally) can be represented by means of membership functions over numerical bases (e.g., Hersh & Caramazza, 1976; Zadeh, 1975). Representations of this sort might be useful in formal decision or risk analyses because they provide a mathematical means for handling forms of uncertainty that are not well represented by probability theory (Watson, Weiss, & Donnell, 1979).

Wallsten, Budescu, Rapoport, Zwick, and Forsyth (1985) provide a full discussion of membership functions, a method for empirically deriving them, and a demonstration that such functions can be established in a reliable and valid manner within a specific, well defined context. However, for this approach to risk analysis to have any hope of success, it is necessary that the membership functions for specific expressions remain relatively fixed over individuals and over contexts. Even within the single context of the Wallsten et al. study there were substantial individual differences in the membership functions for a given expression. These results do not indicate, of course, whether an individual's membership function for a particular phrase changes systematically over contexts.

Related research suggests that context is important. A few studies (Cohen, Dearnley, & Hansel, 1958; Borges & Sawyers, 1974) have shown that the interpretations of quantifiers of amount, such as <u>some</u>, <u>several</u>, <u>many</u>, and so on, are affected quantity of the object available, or by properties of the objects involved (Hörmann, 1983). For example, both Borges and Sawyers and Cohen et al. had subjects take <u>a few</u>, <u>some</u>, <u>several</u>, etc., marbles from trays containing differing numbers of marbles. The more marbles there were in the tray, the more that were taken in response to any given request. Thus, the number corresponding to a particular quantifier increased with the total number available.

Similarly, in a review of research on the quantification of frequency expressions, Pepper (1981) concluded that such expressions have a usual meaning as well as a range of meanings that varies with person and context. In particular, the meanings of at least some phrases vary as a function of the usual or expected frequency of the event being described. Pepper's (1981) conclusion rests in part on a study by Pepper and Prytulak (1974) utilizing quantifiers of frequency such as <u>frequently</u> or sometimes. Subjects were asked the meanings of such phrases in contexts of differing expected frequencies, or in the absence of a context. In each case subjects indicated in how many out of every 100 occasions a specified event occurred. The numerical definition of each phrase was considerably less with a low frequency context than for the others, and somewhat greater for the high frequency than for the null context. These effects were substantial. Thus, for example, the numerical value assigned to very often in the context of earthquakes in California was less than that assigned to sometimes in the context of gun play in Hollywood Western movies. Considering the close correspondence between probability and frequency terms, one would predict that the interpretation of probability terms is likely to be related positively to base rates or to perceived prior probabilities.

From another perspective, one might consider the expression of a probability phrase by an expert or knowledgeable person to be diagnostic information. An individual might combine this diagnostic information with his or her prior judgment about the event to yield a revised judgment. However, it has been demonstrated that under a variety of conditions people are insensitive to base rates when processing diagnostic information (Bar Hillel, 1983; see also Wallsten, 1983). Extrapolating from this line of research, base rate should have little or no effect on the interpretations of probability phrases.

Thus, the purposes of this paper are (a) to ask whether, and if so, how, the meanings of proability expressions are influenced by the base rates or expected probabilities of the events they modify, and (b) to replicate and extend the analogous work of Pepper and Prytulak (1974) with frequency expressions. Two experiments are reported. Experiment 1 utilized professional meteorologists as subjects, and demonstrated that even they, who use probability terms regularly to convey levels of uncertainty to the public, interpret such terms as a positive function of the base rates of the events being predicted. Experiment 2 employed college students as subjects within a more complete design to explore more fully the parameters of the phenomenon.

Experiment 1

Meteorologists were asked to interpret verbal expressions of uncertainty in medica contexts. Meteorologists were selected as the subjects for two reasons. First, the clear communication of uncertainty is important to them. They issue probabilistic forecasts on a regular basis, and they frequently do so with nonnumerical probability phrases. Second, in the context of the probability of precipitation (POP), the National Weather Service (NSW) has actually assigned certain probabilities to specific phrases (National Weather Service 1984, Chapter C-11). If terms that are given probability assignments and are used on a day to day basis in one context are, nevertheless, influenced by base rate considerations in another, then the importance and pervasiveness of the effect is clearly established.

It is important to understand how verbal expressions are used in POP forecasts. The one weather event for which numerical probabilistic forecasts are provided to the U.S. public is that of precipitation. In the case of precipitation, the National Weather Service prescribes that the NWS forecaster must provide a numerical probability POP judgment, and then may, at his or her option also express this judgment nonnumerically. If the forecaster chooses to use a nonnumerical probability phrase, then a probability of 0.10 or 0.20 must be translated as <u>slight chance</u>, 0.30, 0.40, or 0.50 as <u>chance</u>, and 0.60 and 0.70 as <u>likely</u>. Other probability terms are not allowed for POP forecasts, but they can be used in other ways. For example, <u>possibly</u> might be used in a forecast such as "a chance of rain today, possibly heavy at times." Non-NWS forecasters (e.g., TV weather forecasters) are not bound by these rules, but are generally aware of them.

Thus, an experiment was designed to answer two questions. First, would the base rate frequencies of medical events affect meteorologists' probability interpretations of the probabilistic modifiers that they use regularly in weather forecasting? Second, would meteorologists interpret probability phrases in a medical situation according to values they have been instructed to use or are aware of in precipitation forecasting?

A pilot study was run involving 20 NWS meteorologists. On this basis a more complete study was undertaken with a larger sample.

Method

<u>Subjects</u>. Questionnaires were sent to 60 meteorologists, including NWS forecasters, television forecasters, and research meteorologists, who were members of a local chapter of the American Meteorological Society. The cover letter promised that their responses would be discussed at a forthcoming meeting of their group and indicated that the experimental results might be

published.

Questionnaire and design. A sample questionnaire is shown in Table 1. Note that the first and third contexts, which can be referred to as the coffee and ankle contexts, respectively, both represent high probability events. Contexts 2 and 4, referring to wart and flu situations, respectively, represent low probability events. High and low probability contexts were selected informally following discussion with a medical consultant. Note also the use of four probability phrases, likely, possible, chance, and slight chance. These terms were selected because they are commonly used in weather forecasts and because three of the four terms have been assigned meanings by the NWS in the context of POP forecasts.

The four basic contexts were combined with the four probability phrases in two different 2 x 2 designs as shown in Table 2. Half the meteorologists received the four context-probability phrase combinations defined by the major diagonal of the first 2 x 2 design (likely-coffee and possible-wart), and the minor diagonal of the second 2 x 2 design (chance-flu and slight chance-ankle). The other half of the meteorologists received the remaining four combinations. Thus, each meteorologist received each scenario and each probability phrase once, but factorial designs were achieved that are necessary for suitable statistical analyses.

Subjects were instructed that they could respond with either a single probability or a probability range. Responses were returned by mail.

Table 1

Sample Questionnaire for Experiment 1

You normally drink about 10-12 cups of strong coffee a day. The doctor tells you that if you eliminate caffeine it is likely your gastric disturbances will stop.

What is the probability that your gastric disturbances will stop?____________

You have a wart removed from your hand. The doctor tells you it is possible it will grow back again within three months.

What is the probability it will grow back again within three months?____________

You severely twist your ankle in a game of soccer. The doctor tells you there is a slight chance it is badly sprained rather than broken, but that the treatment and prognosis is the same in either case.

What is the probability it is sprained?____________

You are considering a flu shot to protect against Type A influenza. The doctor tells you there is a chance of severe, life threatening side effects.

What is the probability of severe, life threatening side effects?____________

Results

Forty-six responses were received, for a return rate of 77%. Of the 184 probability estimates (46 subjects x 4 estimates/subject), 20% were given as the probability ranges and the rest as single numbers. The range estimates were roughly equally distributed among the four phrases. The subsequent analyses utilized the point estimates plus the midpoints of the probability intervals.

Figure 1 shows stem and leaf plots of the probability estimates in each of the eight cells of the design. The variability is considerable. Furthermore, although the response distributions cover the NWS-assigned values for slight chance, chance, and likely in all cases, in only three of the six instances are these values at the modes (slight chance-ankle, chance-ankle, and likely-wart).

Table 2 shows the mean estimate in each condition. It is clear that on the average a given expression was interpreted as reflecting a higher probability when it was used to predict the high base rate than the low base rate event.

The impression from Table 2 is confirmed by statistical analyses performed separately on the two matrices in the table. Within each matrix, one group of subjects responded in cells (1,1) and (2,2), while the other group responded in cells (1,2) and (2,1). Thus, the main effect of context was tested by first assigning a score to each subject for each matrix equal to the difference between his or her two responses in that matrix. A t-test comparing the two groups of difference scores for each

Stem	Likely: Coffee	Likely: Wart	Chance: Ankle	Chance: Flu
.9	0000	000		
.8	00000000055	0005		
.7	0000000	000555	0	
.6		055	00	
.5	0	0	000000	0000
.4		00	00	
.3	0	03	00005	000
.2			005	000
.1			05	0000005
0				00011555

Stem	Possible: Coffee	Possible: Wart	Slight Chance: Ankle	Slight Chance: Flu
.9				
.8				
.7	05	0		
.6	00000		0	
.5	0000000	00000000000	00	
.4	05	00		
.3	0003	00003	000	0
.2		0005	000000000000	0000
.1	0	0	0000055	00055
0		1		0011355555

Figure 1. Stem and leaf plots for Experiment 1, showing response distributions in each of the eight cells of the design.

Table 2

Experimental Design and Mean Responses for Experiment 1

	Context	
Phrase	High Probability	Low Probability
	Coffee	Wart
Likely	.75	.67
Possible	.48	.38
	Ankle	Flu
Chance	.39	.18
Slight change	.23	.10

matrix tested the null hypothesis that $\mu_{11} - \mu_{22} = \mu_{12} - \mu_{21}$. If base rates positively affect probability estimates, then $\mu_{11} - \mu_{22} > \mu_{12} - \mu_{21}$. The *t*-tests were highly significant in both cases, with $t(44) = 3.17$ and 5.99 for the top and bottom matrices, respectively.

The phrase-context interaction was tested by first assigning a score to each subject for each matrix equal to the sum of his or her two responses in that matrix. A *t*-test comparing the two groups of sum scores for each matrix tested the null hypothesis that $\mu_{11} + \mu_{22} = \mu_{12} + \mu_{21}$. The results were nonsignificant in both cases, with $t(44) = -0.27$ and 1.56 for the top and bottom matrices, respectively. Thus, in each matrix, the effect of context was not significantly different for either of the two phrases.

Discussion

Two results are clear. First, in this medical context the meteorologists were not particularly constrained in interpreting the probabilistic phrases by the numerical conversion mandated by the NWS for precipitation forecasts. Thus, this relatively homogeneous group of subjects was no less variable in converting probability terms to numbers than have been subjects employed in other studies asking for numerical conversion of probability phrases (Budescu & Wallsten, 1985). It should be noted that not all the respondents were NWS forecasters. However, all were interested in forecasting and generally aware of NWS policies. Furthermore, similar results were obtained in the pilot study, which was limited to NWS forecasters.

Second, and directly bearing on the goal of the present work, the meteorologists' interpretations of probability expressions in this medical context varied as a positive function of event base rate. It must be emphasized that nothing in the instructions nor in the questionnaire mentioned base rate or that the various predicted events actually occur with differing relative frequencies. Nevertheless, this variable had a profound effect on the responses of this sophisticated group of subjects, demonstrating the robustness of the phenomenon.

Experiment 2

The purpose of this experiment was to investigate under more controlled circumstances the relation between perceived base rates and the interpretations of probability and frequency expressions. Such information is necessary if we are to develop a theoretical understanding of how judgment is formed, modified, and communicated on the basis of verbal expressions of uncertainty.

A pilot study was first run to develop sets of scenarios with identical semantic content that differ only in perceived base rate or probability. In the main study,the calibrated scenarios were utilized in hypothetical predictions made by experts. The expert's level of certainty in each prediction was communicated by means of either a probability or a frequency expression.

Pilot Study

Method

Subjects. Thirty undergraduates volunteered in partial fulfillment of requirements for the introductory psychology course. All were native speakers of English.

Materials. Fifty-six scenarios were devised, each with three levels of a variable designed to induce low, intermediate, or high probability judgments. For example, one of the scenarios was, "What is the probability of filling every seat in Carmichael Auditorium for a _______?" In this example the variable took on high, intermediate, and low levels, respectively, of "Tar Heel basketball game," "symphony concert," and "circus." The scenarios were of two types: person oriented, of which the previous one is an example, or weather oriented, of which "What is the probability of snowfall in Montreal in (September, November, or March)?" is an example.

Three sets of materials were prepared, each consisting of the 56 scenarios, each at one level. Each scenario appeared in each set at a different level. Assignment of scenario level to set was random, such that each set had approximately equal numbers of low, middle, and high variables.

Ten subjects were assigned to each set of materials. The questions were printed sequentially in booklets in separate random orders for each subject.

Procedure. Four to six subjects were run in a group, each responding independently in a booklet. Subjects were asked to indicate how probable or likely they thought each specific event was by giving decimal numbers ranging from zero to one inclusive. Printed instructions said, "0 means that you think the event would never happen, 0.5 means that you think the event is as likely to happen as not to happen, and 1 means that you think the event would certainly happen. Use intermediate numbers to indicate intermediate probability judgments."

Results

Our sole intention was to select scenarios with variable levels such that mean probability estimates were significantly different in the intended directions. There were an insufficient number of scenarios for which the middle level differed significantly from both the lower and the higher for us to proceed with all three levels. Thus, 36 scenarios were selected for which two sets of responses differed in the anticipated direction by a t score of at least 4. These are shown in the Appendix. The first 12 scenarios are weather oriented and the latter 24 are person oriented. The modifiers under each scenario in the Appendix will be discussed in conjunction with the main study.

The mean estimated probabilities of the high levels of the scenarios range from 0.50 to 0.93, and those of the low levels range from 0.22 to 0.76. The differences between the high and low levels of a scenario range from 0.14 to 0.55, with a mean of 0.30 and a standard deviation of 0.09.

Main Study

The 36 scenarios were developed and scaled so that they could be used in the main study as hypothetical predictions by experts who express their uncertainty verbally rather than numerically. By utilizing both levels of a given scenario with a particular phrase (e.g., likely) and eliciting subjects' interpretations of the expert's subjective probability in each case, it is possible to assess the effect of prior probability, or base rate, on the interpretation, while holding semantic content fixed. A limitation with which we shall have to contend is that the scaled probabilities do not go below 0.22.

The nine probability and nine frequency phrases employed in the predictions are shown in the first columns of Table 3. Note that four of each type are toward the higher end of the certainty scale, one of each type is roughly neutral (possible and sometimes), and four of each type are toward the lower end of the scale. Because the meanings of such expressions are not precise (Wallsten, et al., 1985), subjects were asked what probability the expert most likely had in mind, as well as lower and upper bounds on the range of probabilities the expert might have been considering.

Table 3

Scenario and High/Low Effects within Expressions for Experiment 2

	Scenario Effects			High/Low Effects			
Expression	χ^2 [a]	df	r [b]	χ^2 [a]	df	Best high-Best low	Pilot high-Pilot low
				Probability			
Sure			.44*			.106	.259
Likely			.81**			.187	.318
	41.0**	8		>69.5**	8		
Probable			.74**			.142	.297
Good chance			.78**			.223	.345
Possible	>18.4**	2	.71**	13.8**	2	.122	.311
Poor chance			.55**			.066	.312
Unlikely			.42*			.064	.259
	34.9*	8		18.2*	8		
Improbable			.31			.028	.258
Doubtful			.09			.078	.271
				Frequency			
Common			.78**			.163	.300
Usually			.66**			.116	.262
	28.3**	8		>67.3**	8		
Frequently			.69**			.144	.344
Often			.79**			.128	.306
Sometimes	>18.4**	2	.65**	>18.4**	2	.161	.308
Unusual			.30			.038	.294
Seldom			.19			.039	.276
	32.3**	8		12.1	8		
Rarely			-.17			.014	.297
Uncommon			.10			.048	.284

[a] See text footnote 1.

[b] Significance tests are not exactly appropriate here

* $p < .05$

** $p < .01$

Method

Subjects. Seventy-two undergraduate students responded to notices around campus promising a $3 payment for participation in a 30 to 45 minute computer controlled experiment on the meanings of probability expressions. All were native speakers of English. Subjects were randomly assigned to 12 experimental groups, with 6 subjects per group.

Materials and design. Hypothetical expert predictions were developed by combining each of the 36 scenarios in the Appendix with six of the probability or frequency expressions shown in Table 3. The expressions assigned to each scenario are shown below each one in the Appendix. Expressions were not assigned randomly to scenarios, but rather were selected subject to certain constraints yielding 12 sets of predictions made by experts. The number preceding each expression in the Appendix refers to the prediction set number of which it was a part.

One constraint was that extreme expressions not be paired with events whose judged probabilities were extreme in the other direction. Thus an attempt was made to keep all predictions well within limits of believability.

A second constraint was that each of the 12 sets of predictions employ 18 scenarios, while each scenario appear with six expressions. Further, each of the 18 scenarios appeared in a given prediction set at both its high and low level, yielding a total of 36 distinct predictions in each set. Within each prediction set both members of each scenario pair appeared with the same probability or frequency expression. Thus, each expression appeared twice in each prediction set,once at each level of a particular scenario. Expressions were assigned to scenario pairs such that over the 12 prediction sets each expression was utilized with both weather and person scenarios and with scenarios that covered a wide range of perceived base rates.

To summarize, the design can be conceptualized in either of two ways, both of which were utilized for analysis. First, each of the 36 scenarios was utilized at both its high and low level with 6 expressions of uncertainty. Thus, within each scenario there is a 6 x 2, expression x high/low level, design, with repeated measures over the second factor. Alternatively, each of the 18 expressions of uncertainty was employed with both the high and low levels of 12 scenarios. Thus, within each expression there is a 12 x 2, scenario by high/low level, design, with repeated measures over the second factor.

Subjects saw the predictions in the form of sentences. Thus, for example, a prediction based on the first scenario in the Appendix is, "There is sure to be higher air pollution in Louisville, Kentucky, than in Charlotte, North Carolina in August." All predictions for a scenario were written such that the sentences were as similar as possible while maintaining good English usage.

Procedure. The experiment was entirely computer controlled. Subjects first read instructions on the screen. The instructions informed them that they were to consider each sentence as it appeared on the screen to be a prediction by a knowledgeable expert about a particular event. Their task was first to indicate the probability the expert most likely had in mind when making the prediction. This was to be followed by an indication of the lowest probability and then the highest probability the expert conceivably had in mind. Because of results of some preliminary pilot work, the instructions emphasized that the judgments were to be of experts' probabilities and not of the strengths with which the expert held his or her opinions. Following the instructions and then at any point throughout the session, the subject was free to ask procedural questions of the experimenter.

Each of the 12 subject groups received a different set of predictions. Predictions were ordered randomly for each subject with the constraint that one member of each of the 18 scenario pairs appeared in the first half of the session and the other member of each of the 18 pairs appeared in the second half.

The screen cleared for each trial. Then the prediction appeared in the form of a sentence at the top of the screen. Below the sentence was the question, "What probability does the expert most likely have in mind?" A line with an arrow centered on it was drawn below the question. The line was anchored on the left with a zero, on the right with a one, and there was an unlabeled tick at the center of the line. The subject used left and right arrows on the keyboard to move the arrow left and right on the line. When the subject had located the arrow to his or her satisfaction, indicating the expert's most likely probability judgment, then the subject pressed the "Enter" key to register that response. A marker appeared at the location of the arrow on the line, and the question changed to, "What is the lowest probability the expert conceivably had in mind?" The subject could position the arrow any place from the left end of the line to his or her previous judgment. Upon registering the lower bound by pressing the "Enter" key, a left bracket appeared at the location of the lower probability, the arrow went back to the position of the first response, and the question changed to, "What is the highest probability the expert conceivably had in mind?" Now the subject was free to locate the response at any point from the right end of the line to the first judgment. Upon registering that upper bound, the screen cleared and a new trial was initiated.

Results

This section is organized as follows: We first look within scenarios to determine whether responses depended on the probability or frequency expression and on the level of the high/low variable. Next are the analyses of major interest, all of which are done within expressions. The first analysis is concerned with whether probability estimates vary with the high/low variable as predicted and with scenario. Subsequent analyses explore the high/low effect and ask whether the scenario effects can be traced to prior probabilities or to semantic factors. Finally, we consider factors that may affect the vagueness, or range of the estimates.

MANOVAs within scenarios. The first questions are whether the present subjects agreed with the pilot subjects on the relative probabilities of the two levels within scenarios, and whether they attended to the various probability and frequency expressions combined with each scenario to yield the predictions. These questions are answered with the aid of a MANOVA on the expression by high/low level, 6 x 2, design, for each of the 36 scenarios. The three dependent variables are the best probability judgment, the lower bound and the upper bound.

Overall, the expression effect is highly significant. Over the 36 scenarios, the multivariate $\underline{F}(15,174)$ ranges from 1.29 to 6.53 with a mean value of 3.41. This and all subsequent multivariate $\underline{F}$s were calculated according to Wilk's criterion. For 31 of the $\underline{F}$ values, $\underline{p} < 0.01$, and for three more $\underline{p} < 0.05$. From another perspective, the $\underline{p}$ values from $\underline{m}$ separate analyses can be combined for an overall significance test by taking $\Sigma - 2 \ln \underline{p}_i$, where $\underline{i} = 1,\ldots,\underline{m}$. This yields a X^2 statistic with $2m$ degrees of freedom (Rosenthal, 1978). Combining $\underline{p}$-values separately over the 12 weather and the 24 person scenarios results in $X^2(24) > 175.9$ and $X^2(48) > 362.6$, respectively, for both of which $\underline{p} < 0.001$.[1] Thus, subjects were sensitive to the different probability or frequency expressions within scenarios.

The high/low effect is also significant overall, although it is not as strong as that for expression. Over the 36 scenarios, the multivariate $\underline{F}(3,63)$ ranges from 0.43 to 9.18 with a mean of 3.05. In 11 cases, $\underline{p} < 0.01$ and in 5 more $\underline{p} < 0.05$. Combining $\underline{p}$ values over the 12 weather and 24 person scenarios results in $X^2(24) = 69.8$ and $X^2(48) = 166.7$, respectively, for both of which $\underline{p} < 0.001$. The mean differences in the best probability

---

[1] In these and some subsequent cases, lower bounds are calculated for the X^2 values, because the exact probabilities were not available whenever $\underline{p} < 0.0001$

estimates to the high and low levels of a scenario are in the correct direction in all but two cases. The mean differences range from -0.048 to 0.265, with a mean of 0.10 and a standard deviation of 0.06. The effect size is similar for the judged lower and upper bounds.

MANOVAs within expressions. Having obtained the necessary effects in the previous analyses, we now ask whether the interpretations of predictions utilizing a particular probability or frequency expression depended on the scenario and on the level of the high/low variable. These questions are answered by performing MANOVAs on the 12 x 2, scenario by high/low level, design within each of the 18 expressions.

Overall, there is a significant effect of scenario. Over the 18 expressions, the values of the multivariate $F(33,380)$ range from 1.11 to 2.89, with a mean of 1.76. In eight cases, $p < 0.01$, and in six more, $p < 0.05$. Because the patterns of results differ somewhat over the low, neutral, and high probability and frequency expressions, it is prudent to aggregate p-values separately within each of the distinct categories. The resulting chi-square values with their associated degrees of freedom are shown in the designated columns under Scenario Effects in Table 3; all are significant at $p < 0.01$.

The effects of the high/low variable are less consistent overall. The values of the multivariate $F(3,129)$ range from 0.42 to 14.88, and differ systematically over type of expression. The chi-squares and degrees of freedom corresponding to aggregated p-values are shown in Table 3 in the indicated columns under

High/Low Effects. Note the highly significant effects for the high and neutral expressions. The effects are much smaller for the low expressions, and fail to reach significance in the low-frequency case (unusual. . .uncommon).

High/low effects within expressions. The high/low effects are a major focus of the study and require further exploration. The magnitudes of high/low effects on the best probability estimates are shown for each expression in Table 3 in the column, Best high- Best low. The effect sizes are similar for the estimated lower and upper bounds, indicating that when this variable was operative, it shifted the entire range of meaning, not just the best value within that range.

The pattern of significance levels indicated by the chi-square statistics are reflected in the relative effect sizes. Note that all effects are in the correct direction, but that those for the positive and neutral expressions range from 0.106 to 0.223, while those for the low expressions are much smaller, and range from 0.014 to 0.078.

The differences in mean pilot probability estimates between the high and low levels of the 12 scenarios used with each expression are shown in the last column of Table 3. They are consistently greater than the effects on the best estimates in this study. Within each expression the magnitudes of these two effects were compared by means of a t-test for dependent observations. The values of $t(11)$ ranged from 2.79 to 8.14 for which $p < 0.01$ in all cases. Similar results obtained for t-tests comparing the pilot effects to those for the lower and upper bounds. Thus, the high/low variable has a less pronounced

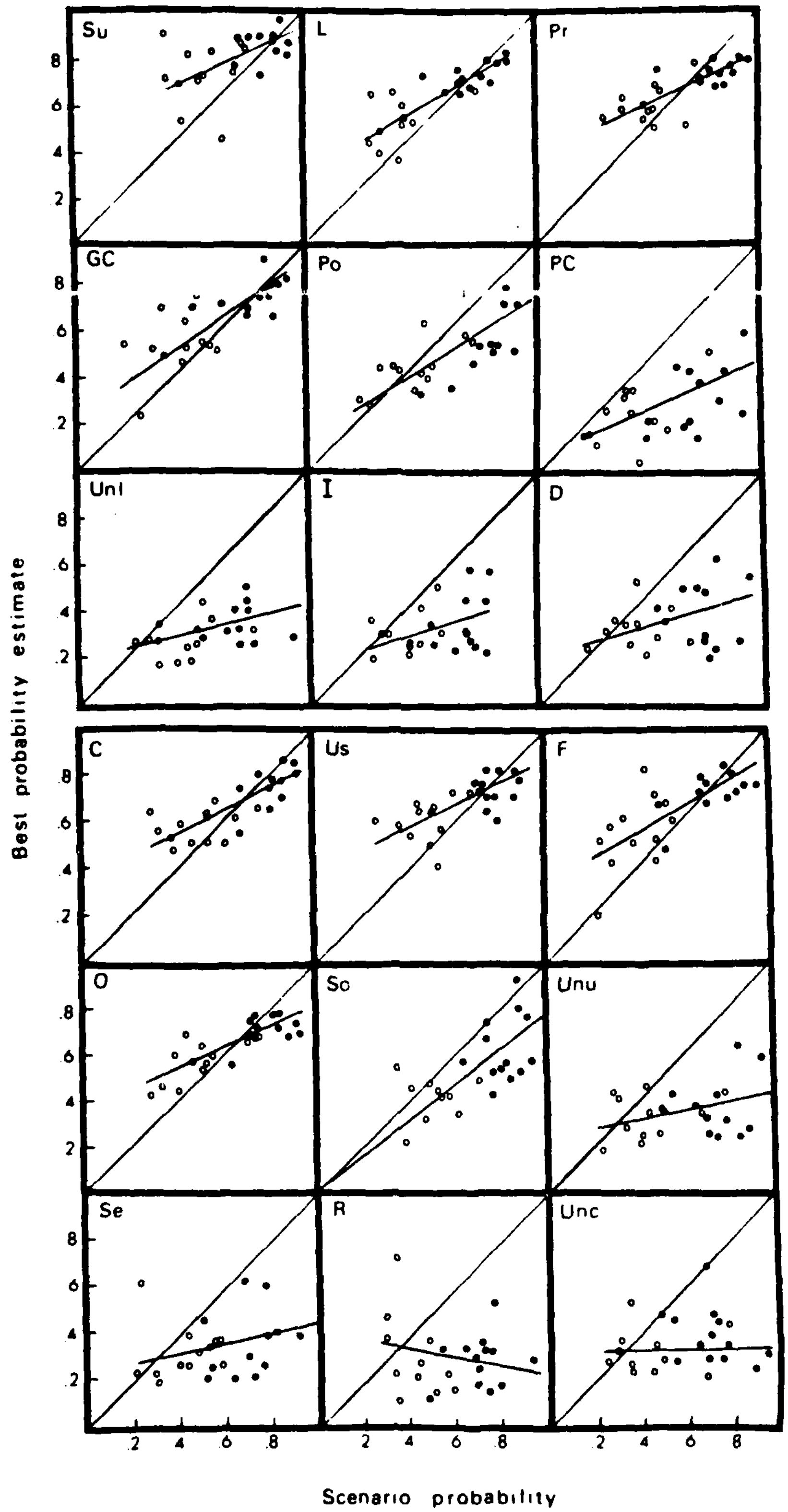

Figure 2. Scatter plots for Experiment 2, showing mean best probability estimates as a function of mean scenario probability as scaled in the pilot study. Closed points represent high scenario level and open points represent low scenario level. The abbreviations are: Su = sure, L = likely, Pr = probable, GC = good chance, Po = possible, PC = poor chance, Unl = unlikely, I = improbable, D = doubtful, C = common, Us = usually, F = frequently, O = often, So = sometimes, Unu = unusual, Se = seldom, R = rarely, Unc = uncommon.

effect in the presence of the probability or frequency expressions than in their absence.

Scenario effects within expressions. The significant high/low effects must be due to differences in scenario probability, because the semantic content is identical for each high/low pair. However, the significant scenario effects may be due in part to differing scenario probabilities and in part to other factors. The role of scenario probabilities in the scenario effects can be seen graphically in Figure 2. The 12 closed dots for each term plot the mean best probability estimates as a function of the scenario probabilities from the pilot study for the high levels of the 12 scenarios used with that term. The 12 open dots plot the mean best probability estimates as a function of the scaled probabilities for the low levels of the scenarios. Thus, each group of subjects contributed two points to each scatter plot, one for the high and one for the low level of a scenario. The correlations between the mean best estimates and the scenario probabilities, ignoring the high/low distinction, are shown in Table 3 in the column labeled r under Scenario Effects. With the exception of six low expressions, all the phrases have correlations of at least 0.44 that are significantly different from zero by the usual test. The significance test is not truly appropriate, however, because each group contributed two points to the correlation, and therefore pairs of points are not independent.

The significant scenario effects for the low expressions are not accompanied by high correlations between the best and scenario probabilities, suggesting that these effects are due to other, perhaps semantic, factors. Of course, although scenario probability clearly plays a role in the other scenario effects, there is no reason to believe that it is the sole factor in those instances.

It is of interest to fit linear functions to the scatter plots in Figure 2. Because there is sampling error in both coordinates of the points, and because our goal is to find the best linear fit rather than simply to predict one set of values given the other, the usual linear regression techniques are not suitable. Rather, the scatter plots were fit with linear structural equations (Isaac, 1969), which simultaneously minimize the sum of squared deviations over both axes. The slopes of these lines are shown in Table 4 in the column labeled $\hat{\beta}$ Standard errors of the slopes are shown in the adjacent column, and _t_ statistics for the hypotheses that $\beta = 0$ and 1 are shown in the next two columns.

Note first that the slopes for the high expressions as well as for _possible_ and _poor chance_ are significantly different from both zero and one. In these cases it is legitimate to conclude that the effect of the phrase is to decrease high scenario probabilities and to increase low scenario probabilities. The point at which the function crosses the diagonal represents the scenario probability that is unchanged by the verbal expression. The diagonal intercepts are shown in Table 4. If it is thought that the subjects' interpretations of the experts' predictions represent some kind of an average between the prior probability of the event and the meaning of the probabilistic modifier, then the diagonal intercept can be taken as the best point interpretation of the meaning of the probability phrase, as indicated in Table 4.

Table 4

Result of Linear Structural Fits to Scatter Plots in Figure 2

Phrase	$\hat{\beta}$	Standard Error of $\hat{\beta}$	t for $\beta = 0$	t for $\beta = 1$	Diagonal Intercept	Mean
Sure	.43	.43	2.82*	2.53*	.88[a]	.80
Likely	.58	.15	32.80*	10.73*	.72[a]	.64
Probable	.43	.19	13.98*	12.56*	.66[a]	.66
Good chance	.67	.17	28.80*	4.87*	.79[a]	.67
Possible	.60	.21	17.65*	5.02*	.39[a]	.53
Poor chance	.40	.33	4.48*	4.94*	.12[a]	.30
Unlikely	.21	.46	0.78	4.03*	.24	.31[a]
Improbable	.27	.66	0.58	1.71*	.25	.33[a]
Doubtful	.25	.63	0.55	1.95*	.28	.36[a]
Common	.51	.17	23.17*	12.19*	.72[a]	.68
Usually	.47	.25	9.87*	6.78*	.73[a]	.69
Frequently	.57	.23	14.52*	5.09*	.76[a]	.65
Often	.45	.16	20.98*	16.47*	.65[a]	.64
Sometimes	.77	.25	15.11*	1.06	.05	.52
Unusual	.20	.68	0.32	1.85*	.28	.34[a]
Seldom	.19	1.07	0.12	0.76	.30	.35[a]
Rarely	-.20	1.20	0.10	0.59	.34	.29[a]
Uncommon	.07	1.92	0.01	0.26	.33	.35[a]

[a]Best interpretation

The slopes for the remaining low probability expressions as well as for unusual are significantly different from one, but not from zero, while those for seldom, rarely, and uncommon are significantly different from neither one nor zero. It is notable in the last three cases that the standard eror of the slope is considerably larger than for any of the other expressions. Inspection of the particular outlying scenarios that led to the extreme standard errors provided us with no insight as to unique meanings the phrases may have been assuming in those instances.

In any case, the expressions with slopes not significantly different from zero are all the low ones, except poor chance, and are those with the generally smallest high/low effects. It is as if these phrases have relatively fixed interpretations that are not influenced by the scenario probabilities. Their best point interpretations are given by their mean values, as indicated in Table 4. The conclusion that the expressions' interpretations are fixed must be tempered by the fact that these phrases were not used with prior probabilities below 0.20. Had such low probabilities been employed, different conclusions may have emerged.

Finally, the slope for some'imes is significantly different from zero, but not from one. Taken at face value, this result suggests that sometimes has no independent meaning of its own, but is interpreted entirely according to scenario probability. However, the scatter plot for sometimes in Figure 2 suggests otherwise. The anomalous statistical result probably occurred because no scenario probability below 0.33 was used in this instance.

As one test of whether the interpretations of the phrases also depended on the semantic content of the predictions, MANOVASs were run within each phrase to compare the responses to the weather and the person oriented scenarios. Significant effects due to scenario type were found for three of the 18 expressions ($p < 0.05$), but this is well within the limits of chance. Therefore, overall, it cannot be concluded that there was an effect due to scenario type.

Vagueness of the interpretations. A final set of analyses looks at the range of the probability estimates, where range is defined as the estimated upper bound minus the estimated lower bound. The greater the range given by a subject to a prediction, the more vague is that subject's interpretation of the meaning of the prediction.

Within each expression, a scenario by high/low, 12 x 2, ANOVA was performed on the range. As was done previously, the p-values from the separate tests were aggregated within expression type. The results are displayed in Table 5. The scenario effect was significant in all cases, except for the neutral frequency term, sometimes. The high/low variable had no effect on the range.

Correlational analyses between range and scenario probability and between range and best estimated probability do not indicate any systematic relations. However, the magnitude of the range is negatively related to the distance of the best probability estimate from 0.5. Over the 18 expressions, this correlation ranges from -0.03 to -0.66, with a mean value based on r to z transformations of -0.36.

Table 5

χ^2 Values for Scenario and High/Low Effects on Range in Experiment 2

	df	Scenario	High/Low
		Probability Expressions	
High	8	29.0*	9.2
Neutral	2	10.2*	0.7
Low	8	25.8*	4.1
		Frequency Expressions	
High	8	20.1*	11.7
Neutral	2	2.3	0.6
Low	8	21.1*	2.0

*$p < .01$

In addition, the ranges for the different types of expressions (high, positive and low) differed systematically. The mean range for the neutral expressions is 0.30, that for the positive expressions is 0.25, and that for the low expressions is 0.23. In testing these differences statistically, it is necessary to take into account the differential range effects due to scenario. Recall that each scenario was utilized with six expressions (c.f. the Appendix). All three expression types were used with some scenarios and only two were used with other scenarios. The mean range for each expression type was calculated within each scenario. In each of the 8 scenarios involving both low and neutral expressions, the range for the neutral exceeds that for the low expression. Similarly, the range for the neutral is greater than that for the high expression in 12 of 18 scenarios. Finally, the high expression range exceeds the low expression range on 18 of 26 occasions.

As a final test of the difference in ranges, $\underline{t}$-tests for dependent observations were calculated that compared the ranges within scenarios between neutral and high expressions and between high and low expressions. Specifically, for each of the 18 scenarios utilizing both neutral and high expressions, a difference score was calculated equal to the mean range for the neutral expressions minus the mean range for the high expressions, and a $\underline{t}$-statistic was calculated asking whether the mean of these 18 difference scores deviated significantly from 0. The result is $\underline{t}(17) = 2.92$ ($\underline{p} < 0.01$). Similarly, the $\underline{t}$-test for the difference between high and low expressions yields $\underline{t}(25) = 2.26$ ($\underline{p} < 0.01$). The conclusion is therefore firmly established that the neutral expressions are most vague, followed in order by the high and low expressions.

General Discussion

The two experiments taken together provide a strong demonstration that the interpretation of nonnumerical probability or frequency expressions generally depends on the base rate, or prior probability, of the event being described. Experiment 1 indicates the pervasiveness of the phenomenon. Meteorologists, for whom the communication of uncertainty is important, interpreted verbal probability predictions in a medical context in a manner that depended on the base rates of the events, despite the fact that three of the four probabilistic expressions had specified numerical meanings in their professional work. It must be added that the subjects knew when filling out the questionnaire that their collective responses would be discussed at a forthcoming meeting of their association. Therefore, it can be assumed that they were motivated to provide their best judgments. Clearly, if members of this group demonstrate a base rate effect, then most other people will as well.

Experiment 2 utilized college undergradutes in a more standard experimental setting and yielded base rate effects of approximately the same magnitude as were obtained in Experiment 1. The results of Experiment 2 provide some insight into the nature of the phenomenon. They suggest a theoretical explanation and raise a question for which we do not currently have a good answer.

Before focusing on these issues, it is important to discuss our manipulation of base rates. In neither experiment was the concept of base rates, or prior probabilities, mentioned to the subjects, nor were the subjects' base rate judgments obtained. This feature has two implications. First, in view of the considerable individual differences in the judgments of probabilities and the interpretation of probability phrases (Budescu & Wallsten, 1985; Wallsten et al., 1985), the exact parameter estimates obtained in this study should not be taken too seriously. Second, despite the subtlety of the base rate manipulation, it was very effective. This fact must be contrasted with the large number of studies showing people to be relatively insensitive to base rates when making judgments on the basis of diagnostic information (Bar Hillel, 1983; Kahneman &

Tversky, 1973; Tversky & Kahneman, 1982). As Bar Hillel (1983) correctly pointed out, the question is not, why do people ignore base rates, but rather, under what conditions do they utilize them? Based on a thorough literature review, she suggested that base rates are utilized when they are perceived as being causally related to the event in question, when they are stated in a specific manner, and when they are especially concrete or vivid. None of the three conditions was met in the present study.

The present experiments differ from all the others on the use of base rates, in that the others presented subjects with explicitly diagnostic information, whereas we gave them verbal predictions from experts. These two types of information differ in many ways, any of which might be important in determining the weight given to base rate information. It must be emphasized that subjects were not attending to base rates simply because they found the experts' predictions useless, since judgments depended on the latter as well.

Turning now to the present data, it is noteworthy that the probability and frequency expressions yielded very parallel results. In particular, there are two facts for which any theory of the base rate phenomenon must account. The first is the systematic differences in the nature of the high, neutral, and low expressions. The neutral phrases are the most vague in their meaning, while the low phrases are the most precise. Similarly, the interpretations of the high and neutral terms are strongly affected by base rate, whereas those of the low phrases are affected very little and possibly not at all.

It must be emphasized that the difference between the low probability or frequency phrases and the others is not an artifact. Wallsten et al. (1985) empirically established individual subject membership functions for a variety of probability phrases. Although they did not note it in their paper, it is the case that the functions for the high and neutral phrases tend to indicate greater vagueness than do the functions for the low phrases. In addition, differences across subjects in the meanings of the expressions are less for the low than for the high or neutral ones. Furthermore, Borges and Sawyers (1974), Cohen et al. (1958), and Pepper and Prytulak (1974) all found that the lower quantifiers were less sensitive to expected frequency or to background quantity than were the other quantifiers. In fact, Pepper and Prytulak (1974) expected such a result, writing that "in natural language, higher frequency expressions appear more flexible in definition than lower frequency expressions" (p. 96).

The second result for which a theory must account is that the phrases did not simply have an additive effect on the prior or scenario probabilities. Rather, the high terms and possible increased the lower scenario probabilities and decreased the higher scenario probabilities, with the points separating the higher and lower probabilities (the diagonal intercepts in Figure 2) increasing from possible to sure or almost certain. Thus, the meanings of the verbal expressions and the scenario probabilities were being combined by the subjects in some sort of an averaging rather than an adding manner. (As remarked earlier, a similar result might have been apparent with sometimes if that expression had been combined with lower scenario probabilities.)

One way to understand the present results is to assume that a probability phrase has a relatively fixed, but vague core meaning for an individual, perhaps such as can be represented by a membership function over the [0,1] interval. In addition, the individual has a vague judgment of the probability of the event in question, which, perhaps might also be represented as a function over the [0,1] interval. Upon receiving a verbal probabilistic prediction about the event, the person interprets that prediction as a weighted average of two vague probabilities, that which he or she associates with the expression, and that which he or she associates with the event. The weight given to the scenario probabilities might depend on how much independent information or knowledge the individual has about the event in question, although our data do not speak to that issue. Clearly, however, low probability expressions are given more weight in the averaging process than are neutral or high probability expressions. As slight chance in Experiment 1 and poor chance in Experiment 2 demonstrate, low expressions do not always dominate the averaging process. The question for which we do not have a good answer is why low expressions should be given so much weight in general. Pepper and Prytulak (1974) suggest that high frequency expressions are more flexible than are low frequency expressions, and, therefore, of course, they would be given less weight. But this explanation still begs the question as to why that should be the case, which remains an important, unresolved issue to which we are directing some current work.

References

Bar-Hillel, M. (1983). The base rate fallacy controversy. In R. W. Scholz (Ed.), Decision making under uncertainty. Amsterdam: North Holland.

Borges, M. A., & Sawyers, B. K. (1974). Common verbal quantifiers: Usage and interpretation. Journal of Experimental Psychology, 102, 335-338.

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. Organizational Behavior and Human Decision Processes, in press.

Cohen, J., Dearnley, E. J., & Hansel, C.E.M. (1958). A quantitative study of meaning. British Journal of Educational Psychology, 28, 141-148.

Cole, P., & Morgan, J. L. (Eds.) (1975). Syntax and semantics: Volume 3. Speech acts. New York: Academic Press.

Fries, N. (1980). Ambiguität und Vagheit. Töbingen: Niemeyer.

Hersh, H. M., & Caramazza, A. (1976). A fuzzy set approach to modifiers and vagueness in natural language. Journal of Experimental Psychology: General, 105, 254-276.

Hörmann, H. (1983). How many are some, several, and a few? In R. Baöerle, C. Schwarze, & A. von Stechow (Eds.), Meaning, use and interpretation of language. Berlin/New York: De Gruyter.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.

Kess, J. F., & Hoppe, R. A. (1981). Ambiguity in Psycholinguistics. Amsterdam: Benjamins.

National Weather Service (1984). Weather Service Operations Manual. Silver Springs, MD: National Oceanic and Atmospheric Administration.

Pepper, S. (1981). Problems in the quantification of frequency expressions. In D. Fiske (Ed.), New Directions for Methodology of Social and Behavioral Science (9): Problems with Language Imprecision. San Francisco: Jossey Bass.

Pepper, S., & Prytulak, L. S. (1974). Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions. Journal of Research in Personality, 8, 95-101.

Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85, 185-193.

Searle, J. (1975). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.), Syntax and Semantics: Volume 3. Speech Acts. New York: Academic Press.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rate. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge, England: Cambridge University Press.

Wallsten, T. S. (1983). The theoretical status of judgmental heuristics. In R. W. Scholz (Ed.), Decision making under uncertainty. Amsterdam: North Holland.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1985). Measuring the vague meanings of probability terms. L. L. Thurstone Psychometric Laboratory Report No. 173. Chapel Hill, NC: University of North Carolina.

Watson, S. R., Weiss, J. J., & Donnell, M. (1979). Fuzzy decision analysis. IEEE Transactions on Systems, Man and Cybernetics, SMC-9, 1-9.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-III. Information Sciences, 9, 43-80.

Zimmer, A. C. (1983). Verbal vs. numerical processing of subjective probability. In R. W. Scholz (Ed.), Decision making under uncertainty. Amsterdam: North-Holland.

Appendix

High and Low Levels of Scenarios Plus Probability and Frequency Expressions used in Experiment 2

1. There is higher air pollution in (Louisville, Pittsburgh) than in Charlotte in August.
   3-sure, 11-unlikely, 4-improbable, 2-frequently, 1-unusual, 12-seldom

2. At least (500, 20) people are killed by heat waves in the USA each year.
   8-sure, 11-poor chance, 1-improbable, 12-often, 4-rarely, 9-uncommon

3. There is snow in Chapel Hill in (November, January).
   9-likely, 10-unlikely, 8-improbable, 11-common, 12-rarely, 1-uncommon

4. There is a (1, 12) degree difference in temperature of city and country in spring.
   10-likely, 7-improbable, 5-doubtful, 2-usually, 6-unusual, 9-seldom

5. There is snow fall in Montreal in (November, September).
   12-probable, 5-improbable, 2-doubtful, 10-frequently, 3-seldom, 1-rarely

6. The temperature will hit (90, 110) degrees in Southern California in August.
   2-sure, 11-probable, 7-possible, 12-usually, 10-often, 3-sometimes

7. Snow will accumulate at least (5, 12) inches during the winter in New York City.
   4-sure, 3-likely, 2-possible, 7-usually, 6-often, 5-sometimes

8. The coastal waters of North Carolina are warm enough to swim in comfortably during (August, May).
   7-sure, 4-probable, 3-possible, 8-frequently, 5-often, 12-sometimes

9. There is snow in the North Carolina mountains during (December, October).
   5-probable, 6-good chance, 9-possible, 4-common, 7-frequently, 2-sometimes

10. There is a layer of ice covering small lakes around Chapel Hill in (October, January).
    8-likely, 1-poor chance, 4-unlikely, 6-frequently, 9-rarely, 5-uncommon

11. The first frost in Chapel Hill will occur by the end of (December, October).
    1-probable, 12-good chance, 4-possible, 9-common, 8-usually, 11-sometimes

12. There is snow on the ground during the month of (January, October) in Washington, D.C.
    1-sure, 12-possible, 9-poor chance, 5-common, 2-unusual, 4-seldom

13. The average adult goes to sleep by (12 midnight, 10 p.m.).
    11-likely, 8-poor chance, 9-improbable, 1-frequently, 6-rarely, 4-uncommon

14. The average American adult has (coffee, applejuice) with dinner.
    7-good chance, 12-poor chance, 3-doubtful, 5-frequently, 10-seldom, 2-uncommon

15. The average worker lives within (15, 2) miles of his/her job.
    12-sure, 2-poor chance, 11-doubtful, 3-often, 4-sometimes, 7-unusual

16. The average female will get married before the age of (29, 19).
    6-likely, 2-probable, 1-good chance, 4-usually, 11-frequently, 9-sometimes

17. A person will drop a non-required course after getting an (F, B) on the first exam.
1-likely, 5-good chance, 4-poor chance, 9-frequently, 8-often, 12-unusual

18. A student who cheats on an exam will get caught if (15, 150) people are in class.
5-likely, 2-unlikely, 9-doubtful, 7-often, 10-rarely, 6-uncommon

19. The average student will stay up past (12 midnight, 3 a.m.) during finals week.
6-sure, 5-poor chance, 3-unlikely, 1-common, 4-unusual, 8-uncommon

20. A student with a GPA of (3.8, 2.5) will continue on to graduate or professional school.
7-likely, 10-poor chance, 11-improbable, 12-common, 3-unusual, 2-seldom

21. A student with a (1500, 1050) SAT will obtain a 4.0 average for at least 1 year.
8-probable, 6-possible, 9-unlikely, 1-often, 10-unusual, 11-rarely

22. A student with an (A, C) average in high school is on the Dean's list at least once in college.
9-good chance, 7-unlikely, 10-doubtful, 8-sometimes, 11-unusual, 12-uncommon

23. Every seat in Carmichael Auditorium is filled for a (Tarheel basketball game, circus).
11-sure, 12-likely, 4-good chance, 2-common, 9-usually, 1-sometimes

24. Calculus III will be failed after getting a (D, B) in Calculus I and II.
5-possible, 3-improbable, 1-doubtful, 12-frequently, 8-rarely, 10-uncommon

25. A student will be admitted to law school if he/she has a GPA of (3.0, 2.5) in college.
4-likely, 1-unlikely, 2-improbable- 6-usually, 3-rarely, 11-uncommon

26. A student with an (A, C) average in high school will attend college.
10-sure, 8-good chance, 1-possible, 6-common, 5-usually, 9-often

27. Two students who have been roommates for (3 years, 2 weeks) will be roommates next year.
10-good chance, 12-unlikely, 7-doubtful, 11-usually, 2-rarely, 3-uncommon

28. A paper due in (3 days, 3 weeks) will be started the day after the announcement.
11-possible, 7-poor chance, 6-unlikely, 4-frequently, 9-unusual, 8-seldom

29. A person knows the names of everyone who lives in his/her building of (5, 15) apartments.
9-probable, 10-improbable, 8-doubtful, 7-common, 5-unusual, 6-seldom

30. Someone will order (french fries, onion rings) with a hamburger.
2-likely, 6-probable, 11-good chance, 1-usually, 3-frequently, 10-sometimes

31. An American will use British expressions after living in London (1 week, 1 year).
5-sure, 3-probable, 6-doubtful, 10-common, 8-unusual, 7-uncommon

32. A couple will have at least 1 child after being married for (5, 1) years.
10-probable, 5-unlikely, 12-doubtful, 3-usually, 2-often, 7-seldom

33. An American adult knows (how to drive a car, a foreign language).
7-probable, 2-good chance, 10-possible, 3-common, 11-often, 6-sometimes

34. The average actor will not have an acting job for (3, 9) or more months a year.
9-sure, 8-unlikely, 12-improbable, 10-usually, 7-sometimes, 5-seldom

35. The average person will live in the same (state, house) all his/her life.
8-possible, 3-poor chance, 6-improbable, 4-often, 1-seldom, 5-rarely

36. Someone will know the names of all his/her classmates in a class of (10, 30) people.
3-good chance, 6-poor chance, 4-doubtful, 8-common, 11-seldom, 7-rarely